ASOCIACION ARGENTINA
DE ECONOMIA POLITICA

ANALES | ASOCIACION ARGENTINA DE ECONOMIA POLITICA

LII Reunión Anual

Noviembre de 2017

ISSN 1852-0022
ISBN 978-987-28590-5-3

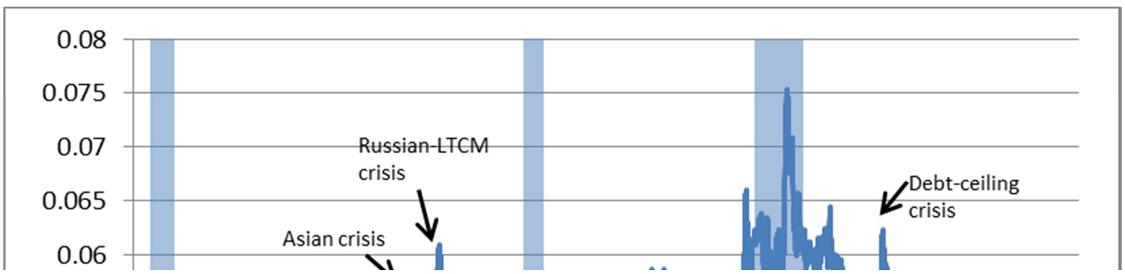
Measuring uncertainty through word vector representations

Aromi, Jose Daniel

3.1 Contemporaneous associations

Figure 1 shows the values of the uncertainty index from 1990 through 2017. An increment in the index can be observed in the three recessions that took place during the sample period. This increment is particularly clear in the case of the recession linked to the 2008 Global Financial Crisis. Additionally, three spikes in the index are observed around three well-known crisis episodes: the Asian Crisis, the Russian Crisis and the 2011 Debt-ceiling crisis. These associations suggest that meaningful information is captured by the index.

Figure 1: Uncertainty index

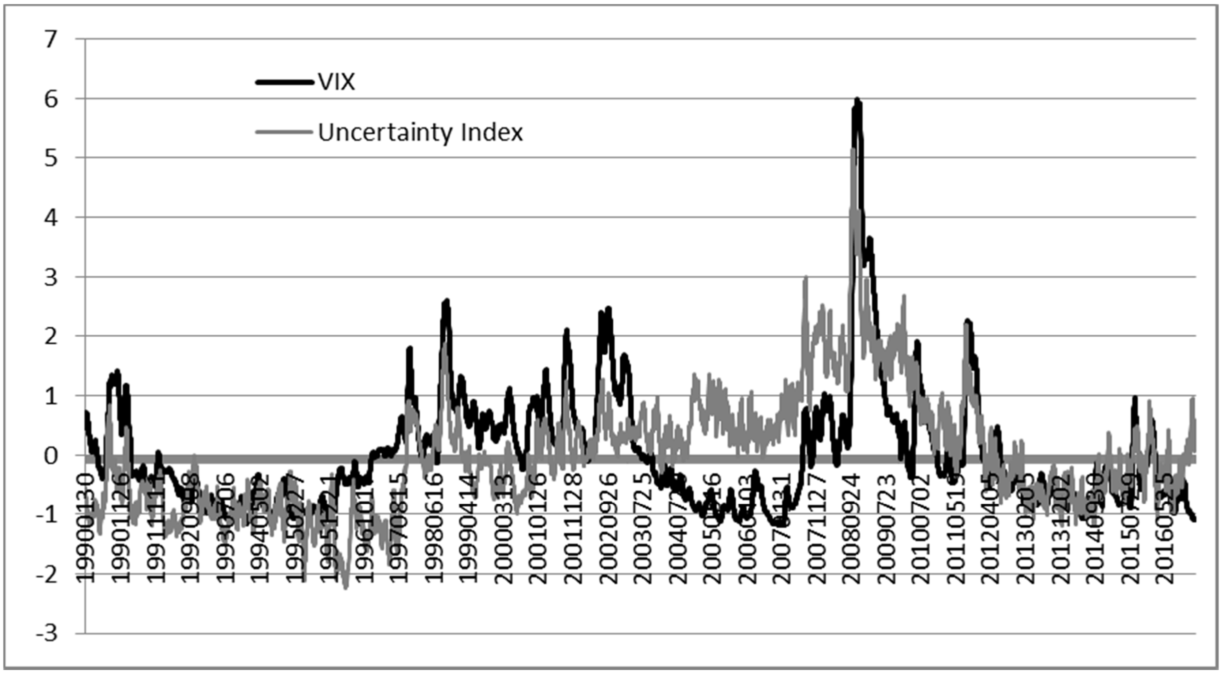


Notes: the figure shows the average value of the index for 20 days moving windows.

In a similar line, figure 2 shows that the uncertainty index is contemporaneously associated to the evolution of expected stock market volatility as summarized by the VIX. The most noticeable co-movement is observed during the 2008 Global Financial Crisis. Other strong co-movements are detected during the recession of the early 1990s and during the Asian and Russian crises. Interestingly, by the end of the sample periods the two indices diverge. This suggests a mismatch between the reaction of market prices and the press to the arrival of Donald Trump to the Oval Office.

Table 2 shows the correlation between the uncertainty index, the VIX and the Equity Market Uncertainty index. Daily series indicate a similar correlation between the VIX and the two other indicators based on press content. The correlation between these two indices based on press content is clearly weaker. Correlations computed for monthly values of the indices confirm the strong association between the VIX and the two other indices. In the case of the uncertainty index, the coefficient of correlation is 0.58.

Figure 2: Uncertainty index vs. VIX



Notes: the figure shows the average value of the indices for 20 days moving windows.

Table 2: Correlation coefficients

A. Daily Values:

	VIX	UI	EMU
VIX	1	0.424	0.093
UI		1	0.381
EMU			1

B. Monthly averages:

	VIX	UI	EMU
VIX	1	0.582	0.127
UI		1	0.556
EMU			1

3.2 Dynamic regressions

Following Corsi (2009) a simple autoregressive model is estimated to describe the dynamic association between VIX values and lagged values of the VIX. This model is later extended to include lagged values of other proxies of volatility. Implementing the heterogeneous autoregressive model (HAR), the value of the VIX on day t is modeled as a function of lagged values for the previous day ($t-1$), average values in the previous week (from $t-1$ through $t-5$), average values in the previous month (from $t-1$ through $t-20$) and a noise term. Formally, the model is given by the following equation:

$$VIX_t = \alpha + \beta_1 VIX_{t-1} + \beta_5 VIX_{[t-5,t-1]} + \beta_{20} VIX_{[t-20,t-1]} + u_t \qquad [1]$$

To evaluate the information content of other uncertainty proxies, this model is later extended to include lagged values of the uncertainty index or the EMU.

Table 3: Heterogeneous Autoregressive Model (HAR)

	[1]	[2]	[3]	[4]	[5]	[6]	[7]
CONSTANT	0.226 ***	0.122	0.200 **	-0.610	0.270 **	-0.484	0.192
	(0.051)	(0.303)	(0.095)	(0.439)	(0.117)	(0.464)	(0.120)
VIX[-1]	0.851 ***	0.851 ***	0.851 ***	0.851 ***	0.851 ***	0.850 ***	0.851 ***
	(0.024)	(0.025)	(0.025)	(0.025)	(0.024)	(.025)	(0.024)
VIX[-5,-1]	0.097 **	0.096 **	0.097 **	0.089 **	0.098 **	0.095 **	0.096 **
	(0.040)	(0.039)	(0.040)	(0.037)	(0.039)	(0.038)	(0.040)
VIX[-20,-1]	0.040 **	0.041 **	0.041 **	0.043 ***	0.040 **	0.038 **	0.040 **
	(0.018)	(0.017)	(0.018)	(0.016)	(0.018)	(0.018)	(0.018)
UI[-1]	-	2.263	-	-	-	-	-
		(5.517)					
EMU[-1]	-	-	0.010	-	-	-	-
			(0.020)				
UI[-5,-1]	-	-	-	18.088 **	-	-	-
				(8.730)			
EMU[-5,-1]	-	-	-	-	-0.015	-	-
					(0.031)		
UI[-20,-1]	-	-	-	-	-	15.458 *	-
						(9.280)	
EMU[-20,-1]	-	-	-	-	-	-	0.012
							(0.032)

Notes: significance levels: "*" 0.10, "**" 0.05 and "***" 0.01. Standard errors (shown in parenthesis) are estimated following Newey & West (1987, 1994).

Table 3 shows the estimated coefficients for these models. Columns 2 and 3 indicate that previous day values of the uncertainty index and the EMU do not add information beyond that provided by VIX's lagged values. Similarly, in the case of average values of the EMU during the previous week or previous month, no additional information is detected. In contrast, columns 4 and 6 show that average values of the uncertainty index during the previous week or previous month add information on subsequent VIX values. The association is positive, that is, higher values of the uncertainty index anticipate higher values for the VIX. Under a plausible interpretation of this result, the uncertainty index serves as a proxy of an unobserved state (perceived uncertainty). This unobserved state affects multiple outcomes observed in future dates, among them expected volatility inferred from option prices.

The information captured in the uncertainty index could be reflected in implied volatility with significant delays. In other words, if derivative markets incorporate information on the unobserved state gradually, the previously documented association between the uncertainty index and one day-ahead VIX values would only represent a fraction of the dynamic association. To evaluate this possibility, exercises under different forecast horizons are implemented through similar dynamic regressions. The associated models are given by:

$$VIX_{t+h-1} = \alpha + \beta_1 VIX_{t-1} + \beta_5 VIX_{[t-5,t-1]} + \beta_{20} VIX_{[t-20,t-1]} + \beta_{UI} UI_{[t-5,t-1]} + u_t$$

Where $h \in \{5,10,20,40,60\}$ is the forecast horizon. The estimated models are shown in table 4. The fitted coefficients associated to the lagged uncertainty index values increase with forecast horizon. This is consistent with the conjectured delayed response of implied volatility. On the other hand, despite the noticeable increment in estimated coefficients, no significant association is observed beyond 10-day horizon forecasts. Additionally, for 5-day ahead forecasts and 10-day ahead forecasts, p-values are relatively high. Based on these estimations, in the case of longer forecast horizons, the evidence on augmented information from lagged uncertainty index values appears suggestive but weak.

Table 4: Forecasting models for alternative forecast horizons

Forecast horizon:	5-day	10-day	20-day	40-day	60-day
CONSTANT	-2.148	-2.758	-3.015	-2.864	-0.889
	(1.790)	(2.718)	(4.172)	(5.541)	(5.265)
VIX[-1]	0.514 ***	0.533 ***	0.493 ***	0.462 ***	0.291 ***
	(0.038)	(0.068)	(0.085)	(0.079)	(0.057)
VIX[-5,-1]	0.356 ***	0.291 ***	0.242 **	0.012	0.008
	(0.079)	(0.096)	(0.114)	(0.161)	(0.122)
VIX[-20,-1]	0.063	0.052	0.054	0.181	0.297 **
	(0.071)	(0.128)	(0.148)	(0.204)	(0.139)
UI[-5,-1]	69.678 *	98.714 *	135.110	181.999	165.870
	(36.678)	(54.794)	(87.106)	(116.342)	(114.872)

Notes: significance levels: "*" 0.10, "**" 0.05 and "***" 0.01. Standard errors (shown in parenthesis) are estimated following Newey & West (1987, 1994).

One possible explanation for this weak result is linked to a scenario in which the value of information provided by lagged values of the uncertainty index is not constant. In particular, it is likely that, in times of high volatility, asset markets are unable to incorporate the intense flow of incoming information.[4] If this is the case, in high volatility scenarios, a bottleneck in information processing capacity could lead to predictability and to prominently delayed responses.

To evaluate this hypothesis and to shed more light into the information value of the uncertainty index, a more flexible specification is considered. Historic values of the VIX are used to split the sample between high historic volatility days and low historic volatility days. If the previous week VIX values are below average ($VIX_{[t-5,t-1]} < 19.6$) day t is classified as a low historic volatility day (L) otherwise the day is classified as a high historic volatility day (H). Two dummy variables are used to signal low volatility and high volatility days (I_t^L and I_t^H respectively). The flexible model is given by:

$$VIX_{t+h-1} = \alpha + \beta_1^L I_t^L VIX_{t-1} + \beta_5^L I_t^L VIX_{[t-5,t-1]} + \beta_{20}^L I_t^L VIX_{[t-20,t-1]} + \beta_{UI}^L I_t^L UI_{[t-5,t-1]} +$$

$$+\beta_1^H I_t^H VIX_{t-1} + \beta_5^H I_t^H VIX_{[t-5,t-1]} + \beta_{20}^H I_t^H VIX_{[t-20,t-1]} + \beta_{UI}^H I_t^H UI_{[t-5,t-1]} + u_t$$

Table 5 shows the estimated parameters of interest. The evidence on informational gains is stronger than in the case of single regime models. For 5-day and 10-day forecast horizons, lagged values of the uncertainty index are shown to provide information in times of low and high historic volatility. For all

[4] A formal argument in this direction could be based on the idea of limited capacity to incorporate new information as proposed in Sims (2003).

forecast horizons considered, the estimated coefficients are larger in the case of high volatility periods. In the case of longer forecast horizons, the estimated coefficients of the lagged uncertainty index are significant only in the case of high historic volatility periods. Also, the estimated coefficients linked to high volatility period increase with every increment in forecast horizon. This evidence points to more than one regime regulating the information values of the uncertainty index. Additionally, it helps to explain the inconclusive results associated to the model in which a single regime is allowed for.

Table 5: Flexible forecasting models – Selected estimated coefficients

Forecast horizon:	5-day	10-day	20-day	40-day	60-day
$\widehat{\beta}^{L}_{UI}$	58.75 *	80.23 *	98.88	104.45	66.08
	(32.83)	(4.74)	(72.46)	(90.28)	(82.53)
$\widehat{\beta}^{H}_{UI}$	85.97 **	125,97 *	187.62 *	320.37 **	348.56 **
	(42.60)	(6.79)	(112.05)	(150.68)	(152.73)

Notes: significance levels: "*" 0.10, "**" 0.05 and "***" 0.01. Standard errors (shown in parenthesis) are estimated following Newey & West (1987, 1994).

3.4 Out of sample forecasts

In-sample forecasting exercises provide evidence regarding historic dynamic associations. From the point of view of the analyst, since all available information is used, this type of exercise leads to the best prediction (Diebold 2015). On the other hand, these forecasting exercises do not reproduce the forecasting problem as experienced by economic agents. Investors, policy makers, professional forecasters and households need to form expectations based on data available at the time of forecast formulation. To evaluate predictive value from this perspective, out of sample exercises are carried out. Below, two competing models are evaluated in terms of predictive accuracy. The first contender (HAR) is given by the model specified in equation [1]. In the competing model (HAR+UI), the lagged value of the VIX index over the previous month ($VIX_{[t-20,t-1]}$) is replaced by the past values of the uncertainty index. To compute past values of the index, two alternative specifications are considered: average values over the previous week ($UI_{[t-5,t-1]}$) and average values over the previous month ($UI_{[t-20,t-1]}$). These models are estimated using data from rolling windows. More precisely, for one-day-ahead forecasts, data from the preceding 1000 days are used. In the case of longer prediction horizons, the lag of 1000-day windows is increased with forecast horizon in order to avoid forward looking biases.

Table 6: Out of sample prediction accuracy

[A] $UI_{[t-5,t-1]}$

Forecast horizon	RMSE				MAE			
	HAR [A]	HAR+UI [B]	Ratio [B]/[A]	D-M test p-value	HAR [A]	HAR+UI [B]	Ratio [B]/[A]	D-M test p-value
1 day	1.583	1.579	0.997	0.219	0.998	0.995	0.997	0.118
5-day	3.130	3.090	0.987	0.163	2.081	2.063	0.991	0.204
10-day	3.878	3.814	0.983	0.162	2.568	2.527	0.984	0.139
20-day	5.120	4.948	0.966	0.025	3.349	3.203	0.956	0.005
40-day	6.758	6.479	0.959	0.088	4.478	4.226	0.944	0.009
60-day	7.740	7.255	0.937	0.167	5.247	4.829	0.920	0.012

[B] $UI_{[t-20,t-1]}$

Forecast horizon	RMSE				MAE			
	HAR [A]	HAR+UI [B]	Ratio [B]/[A]	D-M test p-value	HAR [A]	HAR+UI [B]	Ratio [B]/[A]	D-M test p-value
1 day	1.590	1.589	0.999	0.781	0.999	0.997	0.998	0.290
5-day	3.119	3.090	0.991	0.264	2.077	2.063	0.993	0.437
10-day	3.894	3.807	0.978	0.062	2.576	2.519	0.978	0.082
20-day	5.151	5.014	0.973	0.085	3.363	3.220	0.957	0.021
40-day	6.752	6.432	0.953	0.131	4.825	4.149	0.860	0.009
60-day	7.769	6.976	0.898	0.148	5.267	4.532	0.860	0.006

Table 6 presents estimations of prediction accuracy for the two models considering six different forecast horizons. Two metrics are considered: root mean square errors (RMSE) and mean absolute errors (MAE). Beyond the descriptive statistics, statistical differences in forecast accuracy are evaluated using Diebold-Mariano tests. The implemented tests are two-tailed and the corresponding adjustments for different forecasts horizons are contemplated.[5]

For short forecast horizons, the metrics of accuracy are very similar and no significant difference in accuracy is detected. Starting with 10-day horizons in the case panel [A] and 5-day horizons in the case of panel [B], the difference in prediction accuracy becomes increasingly noticeable. In terms of statistical significance, the strongest results are observed in the case in which the loss function is the mean absolute error. In the case of root mean square errors, the difference is only significant in the case of intermediate forecast horizons.

Similar exercises in which lagged values for the EMU index are used as predictors indicate no predictive ability for this alternative proxy. This negative result suggests that word vector representations are an important tool for the efficient extraction of information in the press. Word similarity established from

[5] The tests were implemented in platform R using command ‘dm.test” from package “forecast”.

this method allows for the construction of an index that approximates perceived uncertainty more precisely and, in this way, extracts valuable information on future volatility expectations.

4. Concluding Remarks

This work proposes a novel metric on uncertainty that exploits word vector representations as a key input. The index is shown to be closely correlated with alternative proxies for uncertainty. Also, the results suggest that press content can generate valuable information regarding subsequent levels of expected volatility as indicated by option prices. This informational gain is particularly noticeable in times of high volatility.

There are several directions in which this work can be extended. A larger corpus might allow for more informed vector representations and more precise sentiment indicators. More topical indices could be built to capture uncertainty regarding specific economic issues, sectors or regions. Also, beyond horse races, this indicator can be viewed as complementary with other indicators of uncertainty such as implicit volatility from derivative markets, subjective reports from households, professional forecasts and other indicators that use press content. This collection of uncertainty proxies could be used to generate an uncertainty factor that would aggregate these diverse indicators.

5. References

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. (2016)"Measuring economic policy uncertainty." The Quarterly Journal of Economics 131.4: 1593-1636.

Basu, Susanto, and Brent Bundick. (2017) "Uncertainty shocks in a model of effective demand." Econometrica 85.3: 937-958.

Bernanke, Ben (1983). Irreversibility, uncertainty, and cyclical investment. The Quarterly Journal of Economics 98(1), pp. 85–106.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. Journal of Financial Econometrics, 7(2), 174-196

Diebold, F. (2015). Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on theUse and Abuse of Diebold-Mariano Test, Journal of Business and Economic Statistics, Vol 33, 1.

Dixit, Avinash and Robert Pindyck (1994). Investment under uncertainty. Princeton: Princeton University Press.

Bekaert, G., Hoerova, M., & Duca, M. L. (2013). Risk, uncertainty and monetary policy. Journal of Monetary Economics, 60(7), 771-788.

Ilut, Cosmin and Martin Schneider (2011). “Ambiguous business cycles”, NBER WP 17900.

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng. (2015)"Measuring uncertainty." The American Economic Review 105.3: 1177-1216.

Hamid A., and Moritz Heiden (2015) Forecasting volatility with empirical similarity and Google Trends Journal of Economic Behavior & Organization, Volume 117, pp. 62-81.

Leduc, S., & Liu, Z. (2016). Uncertainty shocks are aggregate demand shocks. Journal of Monetary Economics, 82, 20-35.

Newey WK & West KD (1987), A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix. Econometrica, 55,pp. 703-708.

Newey WK & West KD (1994), Automatic Lag Selection in Covariance Matrix Estimation. Review of Economic Studies, 61,pp. 631-653.

Orlik, Anna, and Laura Veldkamp. (2014) “Understanding uncertainty shocks and the role of black swans.” No. w20445. National Bureau of Economic Research.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In EMNLP (Vol. 14, pp 1532-1543).

Sims, C. A. (2003). Implications of rational inattention. Journal of monetary Economics, 50(3), 665-690.